Order effects analysis for metabolomic outcomes

Each study will be analyzed independently, to follow the lead of the other outcomes already analyzed.

For each study:

1. Exclude metabolites that have >80% missing data/below the limit of detection in the dataset that is not already imputed
2. Exclude metabolites that have a CV>30.0%
3. Use imputed data that assigns the minimum detectable value for all missing values
4. Scale to a median of 1
5. Log2 transform
6. Create groups based on order, as described in Christina’s paper. This will be two groups per study (UPF -> MPF, MPF -> UPF, LC ->LF, LF -> LC). **Waiting on guidance or dataset from Juen on how to set up the groups and timepoints.**
7. Run the mixed models on each metabolite independently as univariate models. Below is the SAS code that was used to analyze the other outcomes (Aaron will translate into R). ARM=diet order, so the groups described above; treatment=diet; metabolites=each individual metabolite (one per model).

**Proc** **mixed** data=UrineKB1;

class  ARM Treatment SubjectID;

model METABOLITE= ARM Treatment;

random SubjectID;

estimate ‘ARM LC/LF’ intercept **1** Arm **1** **0**;

estimate ‘ARM LF/LC’ intercept **1** ARM **0** **1**;

estimate ‘LC/LF vs LF/LC’ ARM **1** -**1**;

ods output Estimates = tem;

**quit**;

1. Apply BH correction for multiple comparisons ***within*** sample type (BH correction for plasma, and then independently BH correction for 24-hr urine).
2. Using BH corrected p values, use Fisher’s method for each sample type independently. **Erikka to provide resources and/or code for the Fisher’s Method used in previous paper.**
3. Output will look like Supplementary Tables 6-8 from original paper
   1. Link: https://ars.els-cdn.com/content/image/1-s2.0-S0022316623724085-mmc1.docx
   2. Aaron to think of figures to display the data within the paper